1 Point Estimation

1. Consider a population with the following probability function:

$$f(x|\theta) = \theta(1-\theta)^x, \qquad x = 0, 1, 2, 3, \ldots$$

where $0 < \theta < 1$ and $E(X) = (1-\theta)/\theta$. A sample of size 1000 was drawn from this population, having been observed $\sum_{i=1}^{1000} x_i = 980$.

  a) Obtain the memthod of moments estimate for θ.

  b) Find the maximum likelihood estimator for θ.

  c) Calculate, justifying all steps, the maximum likelihood estimate for the mean of the population.

  d) Reparametrize the distribution as a function of $\mu = E(X)$ and use the new probability function to estimate the population mean.

2. Consider a random variable, say X, whose distribution depends on parameters α and θ , having $E(X) = \alpha\,\theta$ and $Var(X) = \alpha\,\theta^2$. From random sample of size 320 that gave $\sum_{i=1}^{320} = 22.2$ and $\sum_{i=1}^{320} x_i^2 = 535.8$, compute an estimate of the unknown parameters. Justify.

3. The time it takes a student to answer an exam question (in minutes) follows an exponential distribution with parameter λ . A random sample of 40 questions gave a total of 480 minutes.

  a) Obtain the maximum likelihood estimate of λ.

  b) Compute the maximum likelihood estimate for the proportion of questions that are answered in less than 15 minutes.

  c) Consider an exam with 8 questions. What is the probability of answering all questions, given that the maximum time allowed for the exam is 2 hours?

4. Le X be a random variable with probability density function given by

$$f(x|\theta) = \frac{3\sqrt{x}}{2\theta} \exp\left(-\frac{x^{3/2}}{\theta}\right) \qquad x > 0 \quad \text{for} \quad \theta > 0$$

  a) Show that $\hat{\theta} = \dfrac{\sum_{i=1}^{n} X_i^{3/2}}{n}$ is the maximum likelihood estimator for θ.

  b) Knowing that $X^{3/2}$ follows an exponential distribution with mean θ, study the biasedness and consistency of the maximum likelihood estimator for θ.

5. Traffic between 8am and 9am at a certain place was measured counting the number of vehicles that passed at that time. Suppose the counts follow a Poisson process. A random sample of 9 observations was collected, having observed the following number of vehicles:

$$(95, 100, 80, 70, 110, 98, 97, 90, 70)$$

  a) Derive the maximum likelihood estimator for the average number of vehicles that pass by that place between 8am and 9am, and compute the corresponding estimate using the above sample.

  b) Show that the estimator you found is consistent and the most efficient.

  c) Compute the maximum likelihood estimate for the probability that no vehicles pass during at least two minutes.

6. Consider a random sample of size three from an exponential distribution, say (X_1, X_2, X_3). In order to estimate the mean, the following statistics are proposed:

$$T_1 = X_1, \quad T_2 = \frac{X_1 + X_2}{2}, \quad T_3 = \frac{X_1 + 2X_2}{3}, \quad T_4 = \overline{X}$$

a) Prove that all the above estimators are unbiased estimators of the mean.

b) Compute the relative efficiency of the estimators.

7. Let $(X_1, \ldots, X_n)$ be a random sample from a population that follows a binomial distribution with parameters 2 and θ:

$$f(x|\theta) = \binom{2}{x}\theta^x(1-\theta)^{2-x}, \qquad x = 0, 1, 2 \qquad \text{for } 0 < \theta < 1$$

a) Obtain the maximum likelihood estimator for θ and study its biasedness.

b) Knowing that $\mathcal{I}(\theta) = \frac{2}{\theta(1-\theta)}$ show that the obtained estimator is the most efficient for θ.

8. The repair time (minutes) of a certain type of machine, say X, follows a Normal distribution with unknown parameters. A random sample of 10 repair times gave:

$$\sum_{i=1}^{10} x_i = 846 \qquad \text{and} \qquad \sum_{i=1}^{10} x_i^2 = 71607$$

Estimate the probability of the repair time being less than 83 minutes.

9.

10. Consider a population with the following density:

$$f(x|\theta) = 2\theta x^{-3}, \qquad x > \theta, \qquad \theta > 0$$

and a random sample of size five: (15, 8, 10, 5, 17). Knowing that $E(X) = 2\theta$, compute the estimate of θ by the method of moments.

11. The maximum height of the waves (meters) at a certain beach is represented by a random variable with the following density function:

$$f(x|\theta) = \frac{x}{\theta}\exp\left(-\frac{x^2}{2\theta}\right), \qquad x > 0$$

a) Obtain the maximum likelihood estimator for θ from the random sample $(X_1, \ldots, X_n)$.

b) Show that the Fisher information equals $1/\theta^2$. (Note: X^2 follows an exponential distribution with mean equal to 2θ).

c) Show that the above estimator is unbiased, and study its efficiency.

d) Suppose that in the last six years the following maximum heights were observed: $(3.1, 2.4, 2.6, 2.2, 1.9, 2.8)$. Compute an estimate for $P(X > 3)$.

12. Consider a Poisson random variable X describing the number of faults in a piece of fabric from the production of a certain firm. The Poisson parameter $\theta = E(X)$ represents the average number of faults per piece of fabric in the total production of the firm. However, rather than in θ, we can be interested in another quantity, for instance the proportion of faultless pieces of fabric from the total production. Find the expression of such proportion as a function of θ.

13. In a box there are θ balls, numbered from 1 to θ. A random sample of 3 balls was drawn with replacement, having obtained (13, 5, 9). Compute an estimate for θ using the method of moments and the maximum likelihood method, and compare them.

2 Interval Estimation

14. Consider a population X with normal distribution and unknown parameters. From that population a sample of size 25 gave

$$\sum_{i=1}^{25} x_i = 75 \qquad \text{and} \qquad \sum_{i=1}^{25} x_i^2 = 321$$

a) Build the 95% confidence interval for the mean.

b) Build the 95% confidence interval for the standard deviation

15. Let X be a normal population with standard deviation 1.5. Based on a sample of size 25 the following confidence interval was build:

$$(\bar{x} - 0.6162, \bar{x} + 0.6162)$$

a) What is the confidence level of this interval?

b) What should the size of the sample be in order to reduce the interval range to half its size, keeping the level of confidence?

16. In order to compare two teaching methods a class of 22 students was divided in two groups of equal size. Each group was taught a different method (method A and method B) and at the end of the course all the students answered the same exam paper. The results, in a sclae from 0 to 100, were as follows:

$$\begin{array}{ll} \text{Method A:} & \bar{x}_A = 74.8; \quad s_A'^2 = 81.5 \\ \text{Method B:} & \bar{x}_B = 72.1; \quad s_B'^2 = 110.5 \end{array}$$

Consider that the grades, in both groups, follow a normal distribution.

a) Build a 95% confidence interval for the ratio of variances in both groups.

b) Having in mind the previous result, build a 99% confidence interval for the difference of means. Interpret the results.

17. In order to study the market share of a certain product in two cities, say F and G, two samples of size 200 each were collected, having observed, respectively, 140 and 160 consumers of the product.

a) Determine, with a 95% confidence level, the market share in city F.

b) Can we state, with a 99% confidence, that the market share is higher in city F?

18. The battery life (hours) is compared in two mobile phone models, using the following sample information:

		Battery life	
Model	N. mobile phones	Mean	Corrected standard error
A	40	6.5	3.0
B	50	5.5	2.5

Using a 95% confidence interval, investigate whether the average battery life can be considered equal in the two models.

19. The number of faulty pieces produced daily by a certain machine follows a Poisson process with mean equal to two. A new machine was tested, having recorded 40 faulty pieces in 30 days. Supposing that the counts for the new machine follow a Poisson process too, construct a 95% confidence interval for the average number of faulty pieces produced daily by the new machine, and say whether you would consider replacement.

20. The effect of class size on grades is studied using sample information on students attending large and small classes, having obtained the following results:

$$\begin{array}{llll} \text{Large classes:} & n = 100 & \bar{x}_1 = 9.84 & s_1'^2 = 8.5398 \\ \text{Small classes:} & m = 100 & \bar{x}_2 = 10.84 & s_2'^2 = 5.8327 \end{array}$$

a) Construct a 90% confidence interval for the average grade of students attending small classes.

b) Construct a 90% confidence interval for the difference in average grade in the two groups and comment.

c) The total number of passes in the two samples were, respectively, 46 and 66. Using a 90% confidence interval, investigate whether the proportion of passes is higher in smaller classes.

21. The content of one-litre cans of paint follows a Normal distribution with a standard deviation equal to 0.06 litres. From a sample of 16 cans, a sample average of 0.95 litres was observed.

    a) According to the sample information, it was stated that the mean paint content of a can lies between 0.911 and 0.989 litres. Comment and say what is the confidence of such statement.

    b) What would you do in order to reduce to half the width of the previous confidence interval?

22. The amount of fruit contained in the filling of some biscuits follows a Normal distribution with unknown mean and variance. The brand guarantees that, on average, each pack of biscuits contains 40g of fruit. A random sample of 20 packs gave a sample average of 38g and a corrected standard error of 4.8g. In order to test the guarantee of the brand, a confidence interval for the mean was constructed, having obtained (35.75353, 40.24647). Determine the level of confidence associated to the interval.

23. The average free acidity of the olive oil of a certain brand is guaranteed not to exceed 1%. A random sample of 16 bottles gave an average acidity of 1.3%, with a corrected sample variance of 0.16. Considering the free acidity to be well described by a Normal random variable, construct a 95% confidence interval for the average free acidity in order to verify the brand's guarantee.

24. In order to compare the city consumption of two new car models, the MZ-Eco and the W-Green, two random samples of size 9 were collected, having obtained, respectively, 8.6 and 8.2 litres/100km. It is known that for these two models consumption follows a Normal distribution with variance equal to 0.2. Construct a 95% confidence interval for the difference in mean consumption.

25. A farm has two identical tractors, whose fuel consumption per hour follows a Normal distribution with unknown parameters. To control consumption, two random samples were collected of size 9 and 5 respectively, having obtained the following results:

    $$\begin{array}{lll} \text{Tractor 1:} & \bar{x}_1 = 9.63 & s_1^2 = 0.10888 \\ \text{Tractor 2:} & \bar{x}_2 = 9.86 & s_2^2 = 0.0344 \end{array}$$

    Supposing that the variance of consumption can be considered equal for the two tractors, construct a 95% confidence interval for the difference in mean consumption.

26. A TV channel aims at gaining at least a 55% audience share for a new series. A marketing research firm agrees in doing each week telephone interviews to a random sample of 200 customers, recording the number of interviewees watching the new series. A certain week 90 customers were recorded as watching the series. Construct a 95% confidence interval for the proportion of customers following the series that week. Was he target of the TV channel achieved?

27. Let X be a Normal random variable, $X \sim N(\mu, \sigma^2)$, with unknown mean and variance $\sigma^2 = 4$. Compute the minimum and maximum sample sizes such that the width of a 95% confidence interval for the mean is between 2 and 3.